AGARD-CP-579

AGARD-CP-579

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARD CONFERENCE PROCEEDINGS 579

Neurological Limitations of Aircraft Operations: Human Performance Implications

(les Limitations neurologiques des opérations aériennes: les Conséquences pour les performances des équipages)

Papers presented at the Aerospace Medical Panel Symposium held in Köln, Germany from 9-12 October 1995.

NORTH ATLANTIC TREATY ORGANIZATION

Published April 1996

Distribution and Availability on Back Cover

DCIEM 2 95-P-37

Response Latencies to Aircraft Detection Among NORAD Surveillance Operators

R.A. Pigeau[1], R.G. Angus[1], P. O'Neill[2]

[1] Defence and Civil Institute of Environmental Medicine
1133 Sheppard Ave. West, P.O. Box 2000
North York, Ontario M3M 3B9, Canada

[2] Operational Research and Analysis Establishment
Ottawa, Ontario, Canada

A function of North American Aerospace Defence (NORAD) in North Bay, Ontario is to identify all aircraft entering Canadian air space. The first step in performing this task is to detect visually the presence of aircraft from either radar or transponder information presented on display consoles. This challenging, real-world vigilance task was used in two experiments to investigate factors affecting detection latencies. Taken as a whole, the experiments revealed that 1) transponder tracks, with their unique symbology, were more easily detected than radar tracks; 2) performance varied as a function of geographic area of coverage; 3) the midnight shift was particularly sensitive to vigilance decrements; and 4) a 'vigilance decrement' effect can occur in a real world task, but this effect is not as strong as those reported in laboratory studies.

1. INTRODUCTION

Numerous vigilance researchers have questioned the wisdom of generalizing laboratory results to real world tasks [1, 2, 7, 11, 13, 20] Their concerns arise mainly from three observations: 1) laboratory vigilance tasks rarely resemble their real world counterparts; 2) when real world tasks are adequately simulated the results often differ from laboratory findings; and 3) there are "...few, if any, troublesome vigilance decrements in the operational tasks of the real world" [1, p. 737]. While admitting that a large body of important psychological knowledge has accrued from laboratory vigilance research, Wiener [22] also cautions that vigilance must be studied in more complex and operationally valid environments. He notes that real-world vigilance tasks usually involve highly trained and experienced individuals working long hours on critical tasks with relatively low signal rates. Unfortunately, conducting research in such environments, without impeding or disrupting ongoing activity, is often very difficult. Measurement techniques need to be non-intrusive, experimental protocols must be consistent with the demands of the workplace and the results have to be relevant (and helpful) to the individuals begin studied.

We have had the opportunity to study such a real world vigilance task (not a simulation) under daily operational conditions, with control over several factors. One function of North American Aerospace Defence (NORAD) in North Bay, Ontario is to identify all aircraft entering Canadian air space. This is accomplished by: 1) visually detecting potential intruders from either radar or transponder information presented on display consoles; and 2) correlating the detected targets with logged aircraft flight plans. If the targets do not correlate with logged flight data, then interceptor aircraft may be scrambled to visually identify the transgressor. Since the entire process is dependent upon rapid initial detection by the surveillance operators, it is important to have a good understanding of the factors influencing detection performance in this operational environment.

[1] A shorter version of this paper will appear in the Human Factors Journal (Sept '95)

Paper presented at the Aerospace Medical Panel Symposium on "Neurological Limitations of Aircraft Operations: Human Performance Implications", held in Köln, Germany from 9-12 October 1995, and published in CP-579.

1.1 The Air Surveillance Task and Environment

Canadian NORAD Region's air surveillance has been continuously maintained for 35 years and involves monitoring returns from an array of radars covering Canada's east, west and northern territories. The radar information is data-linked to Canadian NORAD Region headquarters in North Bay. Canada's area of radar coverage is divided into two geographic regions, the surveillance of which is given to two separate squadrons (see Figure 1). The squadrons are physically, functionally and operationally independent. They each have their own equipment, personnel schedules and are located in different areas of the NORAD facility. Although they are responsible for different geographic regions, Canada's radar picture is available to either squadron in the event of an emergency.

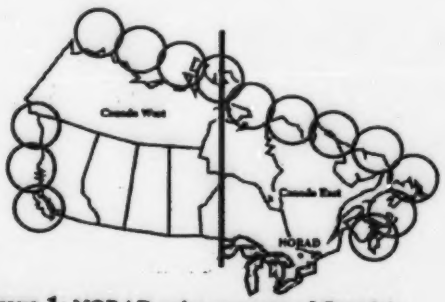

Figure 1: NORAD radar coverage of Canada's geographic perimeter. Neither the number of the radars nor their locations is accurately depicted.

Each geographic region (i.e., Canada East or Canada West) is subdivided into zones that are displayed on individual surveillance consoles. The display consoles are round, monochrome (green phosphor) and vector scanned. Air surveillance operators do not monitor returns from individual radars; they monitor returns from geographic zones that are serviced by multiple radars. The entire air picture is updated every 12 seconds by NORAD computers which must fuse and correlate all radar data to eliminate redundant echoes reflected from adjacent radar sites. To represent aircraft movement, the 7 most recent radar returns are sequentially displayed every 2 seconds. This leaves a 'blip' trail showing an aircraft's heading and speed.

The surveillance operator's task is to detect these tracks and direct the computer to tag and assign track numbers. [NOTE: The operators are not responsible for identifying the tracks, only for detecting them.] Depending on the type of aircraft and its intention, a track can have one of two 'signatures'. Aircraft equipped with transponders emit status and identification signals that are received by the radars and are displayed on NORAD consoles with a 'diamond' symbol. Such tracks are referred to as 'beacon' tracks. Aircraft without transponders can only be detected by search radar. Their symbol, a 'hyphen', is identical to all other spurious radar returns (e.g., noise or clutter caused by storms, sea spray, mountains, etc.) and are more difficult to detect. These tracks are referred to as 'search' tracks.

In order to assess detection performance, a method was needed to measure response latencies of surveillance operators over long periods. This was accomplished by taking advantage of the capability for NORAD's computer system to inject simulated tracks over the live air picture. When displayed on the surveillance operator's console the simulated tracks are indistinguishable from live traffic. Simulated tracks can be given either the 'diamond' (for beacon tracks) or 'hyphen' (for search tracks) symbol and can be allocated precise speeds, headings and altitudes. Both the time a simulated track first appears on the console and the time it is detected by the surveillance operator can be recorded. This allows response times (RT) to be calculated and analysed. Unfortunately, due to inherent system limitations, RTs cannot be computed for live air traffic.

To assess the impact of performing vigilance studies on NORAD operational efficiency, a short pilot study was conducted [4]. During a 48 hr period, 10 simulated tracks per hour were injected into Canada East's (live) airspace. Eighty percent of the 480 simulated tracks were given search track symbols and 20% beacon track symbols. The results indicated that RTs were slower for the detection of search tracks than for beacon tracks. Due to the differences in symbology this result was expected. Also, RT for search tracks was slower for the

midnight shift than for either the day or afternoon shift; RTs for beacon tracks yielded no differences.

The present paper describes two experiments. The first is an extension of the original pilot study [4] investigating the effects of track type (i.e., beacon vs. search) and surveillance shift (i.e., evening vs. midnights) on detection times. The second study varied coverage area and time-on-task. Both experiments were conducted in the Canada East region because of its higher level of air traffic — i.e., the North America - Europe air corridor.

2. EXPERIMENT 1

The initial pilot study [4] demonstrated the feasibility of injecting simulated (aircraft) tracks over live air traffic during day-to-day operations. Experiment 1 was designed to extend these findings by collecting a larger, more representative data set.

2.1 Method

2.1.1 *Subjects*

Sixteen surveillance operators, each with hundreds of hours of logged console time, served as subjects for this experiment. They were informed as to the purpose of the study and were assured that individual performance results would not be released to their superiors. Since the subjects were fulfilling their regular work duties they were not given monetary compensation for participating in the study; however, the subjects were promised a briefing of the final results and they all expressed interest in the study.

2.1.2 *Procedure*

The experiment was conducted on two occasions separated by one month. On each occasion the subjects had recently finished three days of leave and were beginning three consecutive days of either an evening (1500h - 2300h) or midnight (2300h - 0700h) shift as part of their regular duty cycle.

On the first day of each experimental run, prior to the beginning of their shift, the subjects were briefed about the purpose of the study. Although they were informed that simulated tracks would be mixed in with the live traffic, they were given no information concerning frequency or location of these tracks.

For the evening shift the Canada East region was divided into 5 surveillance zones (i.e., consoles) whereas for the midnight shift 4 zones were used. During both shifts the subjects followed a typical work/rest cycle of 45 minutes on scope and 30 minutes rest. The subjects rotated through each console.

Before the experiment, a computer program was written to generate a list of the random stimuli to be manually entered into NORAD's computer during the study. Twenty-five simulated tracks per hour, randomly spread over the 5 (evening shift) or 4 (midnight shift) geographic zones were produced specifying each track's time of appearance, speed, altitude, zone placement and track type (i.e., search or beacon). In total, 2400 simulated tracks were presented (25 tracks/hr x 8hrs/shift x 2 shifts x 6 days) of which 60% were radar and 40% were transponder. Four experienced airforce personnel working in a different area of the NORAD facility used the stimulus list to inject the simulated tracks at the appropriate times.

Since the entire NORAD air picture of detected tracks is continuously recorded on magnetic tapes, we were able to retrieve and generate two result files: 1) the time of appearance for each simulated track and 2) the time of detection for all tracks (simulated and live). We then paired the appearance and detection times for each simulated track and subtracted one from the other to obtain response latencies.

2.2 Results

Forty-six percent (1105) of the 2400 simulated tracks presented to the surveillance operators were lost due to a software problem in the data reduction algorithms — a problem corrected for Experiment 2. Most of the lost data were from days 2 and 3 of the second experimental session. As a result, only 490 (38%) beacon and 805 (62%) search track response latencies were analysed.

Frequency distributions of RT data are typically positively skewed and must be transformed before statistical analyses can be performed. A $\log_{10}$ transformation [19] was applied to the data which resulted in: 1) normally distributed data, and 2) geometric means (i.e., antilogs) in original units of measurement (see Figures 2 and 3). To

remove the influence of outliers in the transformed data, reaction times greater than ± 3 interquartile lengths (i.e., 3 x the range between 25% - 75%) from the mean were excluded from further analysis. This procedure was performed separately for the beacon and search data distributions, and yielded the exclusion of 2 beacon and 4 search scores.

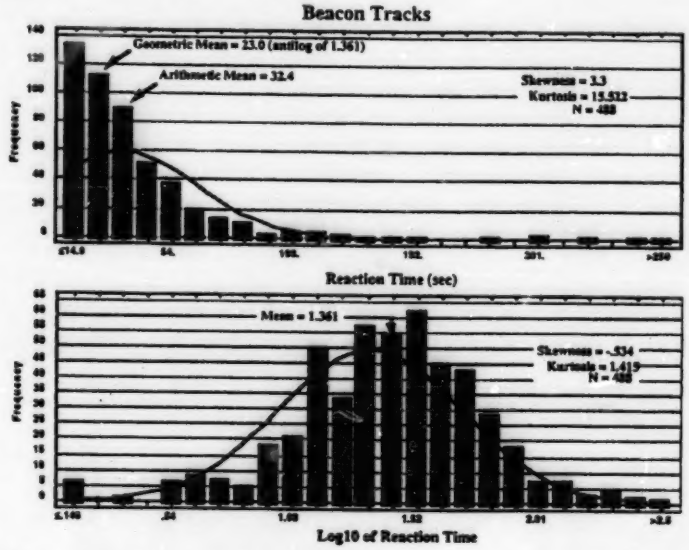

Figure 2: Frequency distributions of detection times to simulated 'Beacon' tracks. Top figure illustrates the positively skewed raw data; bottom figure illustrates the normalized (log10) transformed data.

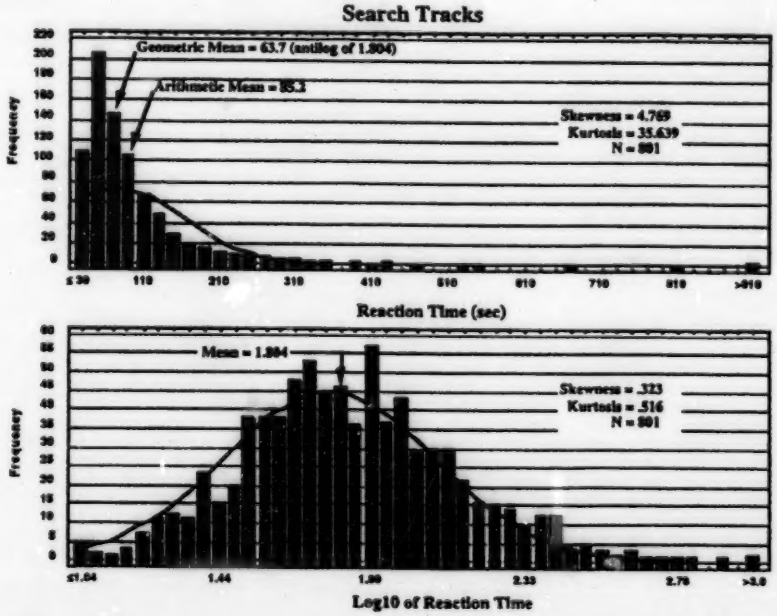

Figure 3: Frequency distributions of detection times to simulated 'Search' tracks. Top figure illustrates the positively skewed raw data; bottom figure illustrates the normalized (log10) transformed data.

A three way ANOVA was performed on the following three factors: track type, shift and hour-into-shift (2 x 2 x 8). Statistically reliable main effects were found for all three factors (track type: $F_{(1,15)} = 497.8$, $P < .0001$; shift: $F_{(1,15)} = 32.73$, $P < .0001$; hour-into-shift: $F_{(7,15)} = 3.55$, $P < .019$) but there were no interactions. Table 1 shows the RT geometric means (with 5-95% confidence intervals) of the main effects for beacon vs. search tracks and the evening vs. midnight shift. RTs were roughly 3 times longer for search tracks (63.7 sec) than beacon (22.9 sec) and the evening shift was 9 seconds faster overall than the midnight shift (48.3 vs. 39.3 sec respectively). Figure 4a illustrates the main effect of RT by hour throughout a shift (collapsed across shift and track type). A marked improvement in performance (≈25 sec) is observed from hour1 to hour2 of a shift. Performance thereafter gradually becomes poorer towards the end of the shift. Although the three way interaction was not significant, Figure 4b suggests that this decline in performance is largely due to slower RTs for search tracks near the end of the midnight shift.

To investigate RTs for the different geographic zones, separate ANOVAs (zone x track type) were performed for the evening and midnight shifts. The evening shift yielded significant main effects for geographic zone ($F_{(4,15)}=5.92$, $P<.0046$) and track type ($F_{(1,15)}=278.75$, $P<.0001$), and an interaction ($F_{(4,15)}=3.78$, $P<.0255$). The midnight shift showed two main effects (zone: $F_{(3,15)}=5.17$, $P<.0119$; track type: $F_{(1,15)}=265.73$, $P<.0001$) and no interaction. Figures 5a & b show the results for the evening and midnight shifts, respectively. During the evening shift search tracks in the North and South East zones are detected much more slowly than in the North East and East. Except for slightly faster RTs in NE1, beacon detection is the same across zones. For the midnight shift, RTs in the North and North East are slower than those in the East, which in turn are slower than those from the South East.

2.3 Discussion

The 30 sec difference in RT between beacon and search tracks is consistent with the findings of the first pilot study. The distinctive symbol (diamond) used for beacon tracks is easily detected on the consoles. Since the radar update interval is 12 seconds the average detection time for beacon tracks is within 2 radar frames (i.e., a trail of 2 'blips'). On the other hand, legitimate search tracks (hyphens) can only be distinguished by their blip-trail trajectories. More time is needed first, to develop a blip-trail history and. second, to detect and differentiate them from background noise (e.g., storms, sea spray). On average, 5 update intervals were needed to detect these tracks (i.e., ≈ 63 sec).

The main effect of shift indicated that on average the evening shift was 9 seconds faster than the midnight shift. It is not clear whether this difference is due to fatigue induced from working the midnight shift or due to the effect of one less console being staffed, compared to the evening shift. The issue of the total number of geographic zones and its overall effect on performance is addressed in Experiment 2.

There were two interesting findings when performance was plotted through time (Figures 4a & b). First, RT was slower during the first hour after a shift change — the effect was most noticeable for search tracks but the trend was also present for beacon data during the first hour of the midnight shift. This result may be due to high levels of social interaction among the operators at the beginning of shifts or it could reflect a latency to re-adopt some unknown 'cognitive set' necessary to efficiently perform the task. Second, there was a progressive increase in RT toward the end of an 8 hour shift.

Performance latencies were also differentially affected by geographic zone. During the evening shift, RTs for radar tracks were longer both in the North and South East zones (see figure 5a). Since the northern zone contained very little air traffic, the slower RTs may have been due to lower operator motivation and arousal. However, the South East zone similarly displayed slower RTs, despite experiencing higher workloads due to evening traffic travelling to

the U.S. from Europe — and thus should have experienced greater arousal and produced better performance. On the other hand, the high amount of air traffic (see Figure 6), may have made detection of search tracks much more difficult.

	Beacon Tracks	*Search Tracks*	**Main Effect for Shift**
Evening Shift	19.5 (17.3 - 21.9) N = 250	58.7 (54.6 - 63.0) N = 436	39.3 (36.5 - 42.3) N = 686
Midnight Shift	27.2 (24.7 - 30.0) N = 238	70.3 (65.6 - 75.2) N = 365	48.3 (45.2 - 51.7) N = 603
Main Effect for Track Type	22.9 (21.2 - 24.8) N = 488	63.7 (60.6 - 67.0) N = 801	**Overall** N = 1289

Table 1: Geometric means and 5-95% confidence intervals of RTs for the shift and track type conditions.

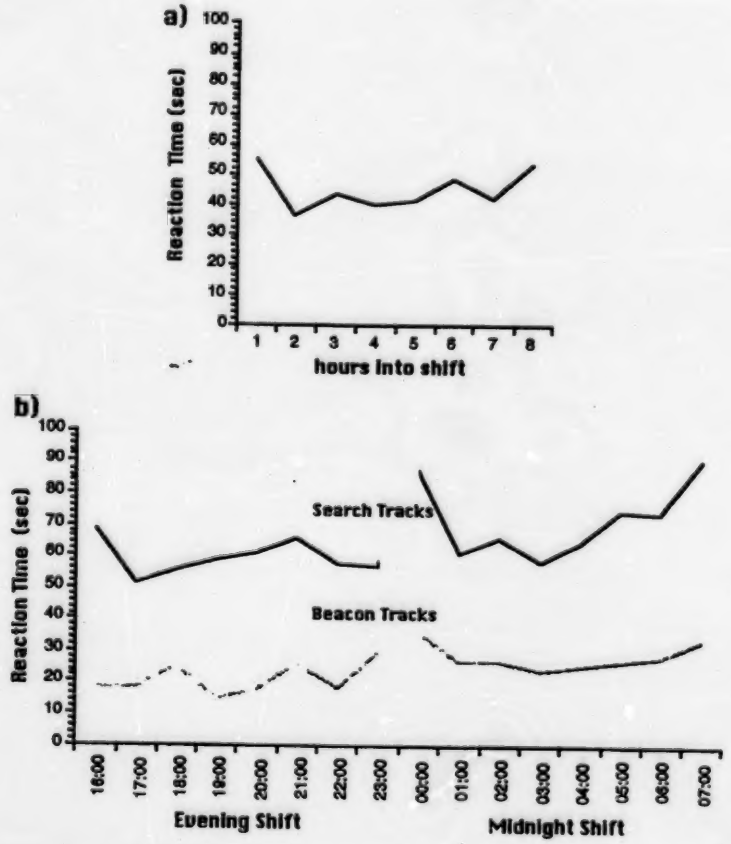

Figure 4: Response times (geometric means) plotted every hour throughout a shift: a) data collapsed across shift and track type; b) data for each shift and track type.

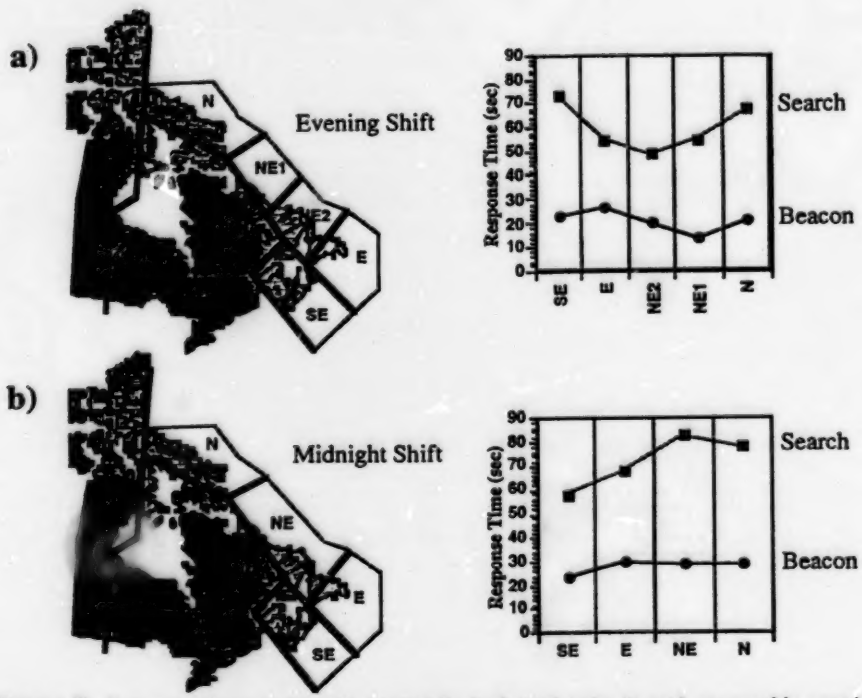

Figure 5: Response times (geometric means) for both track types in each geographic zone: a) evening shift; b) midnight shift.

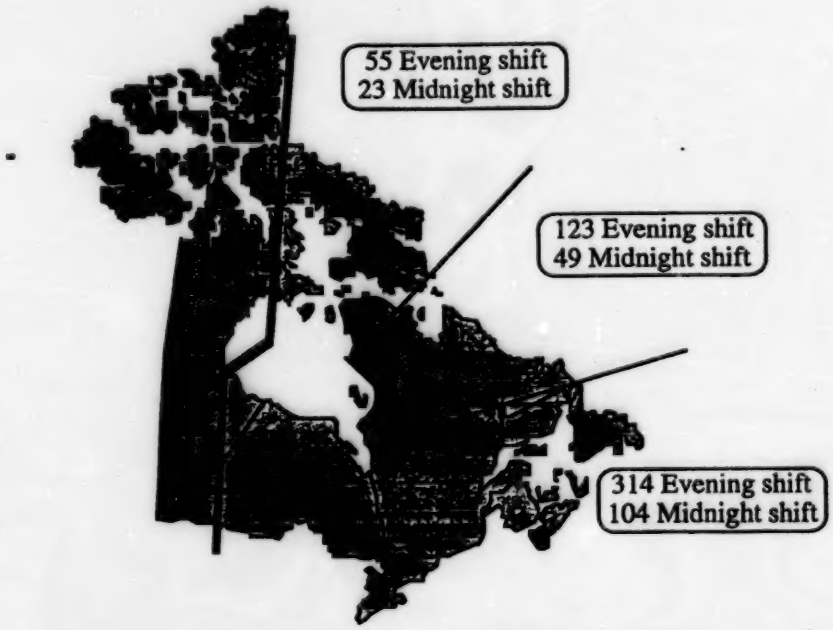

Figure 6: Amount of live air traffic during each shift for different geographic regions.

Figure 6: Amount of live air traffic during each shift for different geographic regions.

During the midnight shift a definite increasing trend was observed for search RTs from the South East to the North (see figure 5b). Slower RTs in the North are likely due to the extremely low amount of air traffic traversing this zone at this time.

In summary, experiment 1 yielded a number of findings: 1) an extended study of a real-world, real-time vigilance task is possible without sacrificing operational efficiency; 2) search tracks are significantly more difficult to detect than beacon tracks, both because they are relatively infrequent events and they have relatively small signal to noise ratios; 3) performance during the first hour of a shift is poorer than during the second hour; 4) detection of search tracks degrades towards the end of a shift — particularly during the midnight shift; and 5) search track performance varies with geographic location of coverage.

3. EXPERIMENT 2

The results from Experiment 1 justified pursuing a more rigorous experimental investigation. Experiment 2 was designed to manipulate two independent variables: coverage area (i.e., number of zones) and time-on-task. The number of geographic zones into which Canada East is divided directly translates into personnel requirements for continuously staffing the consoles. Traditionally, the number of zones usually fluctuated between three and four depending on personnel availability. No data exist demonstrating the degree to which performance is affected by changing area of coverage. Even though the surveillance operator's task is over-learned and extensively practiced, performance could be seriously compromised if the number of geographic zones is reduced; using fewer consoles would require that each console encompass a larger surveillance area. Each aircraft's blip-trail would then subtend a smaller visual angle on the console thus making detection more difficult.

The second independent variable, time-on-task, was included to investigate the vigilance decrement effect observed in Experiment 1, where RTs increased during the latter part of the shift. It was unclear, however, whether performance degraded within individual work periods occurring within any given 8 hour shift. Vigilance research suggests that performance declines should occur after approximately 20 minutes on task [6]. To test this, a condition was added in which the subjects worked for 20 minutes then rested for 20 minutes throughout the duration of a shift, or worked 60 minutes and rested 60 minutes. If the vigilance decrement observed under laboratory conditions can be generalized to real-world tasks then RTs during the 60/60 work/rest condition should be greater than during the 20/20 condition.

3.1 Method

3.1.1 *Subjects*

Sixteen (different) experienced surveillance operators participated in this experiment. Ss were informed as to the general purpose of the study, were assured individual performance would be kept confidential and were promised disclosure of the final results when they became available.

3.1.2 *Procedure*

A 2 x 2 x 2 factorial design was employed to investigate the relationship between coverage, time-on-task and shift. Canada East was divided into either 2 or 4 geographic zones and either 20 min/20 min or 60 min/60 min work/rest cycles were used. Both the evening and midnight shifts received each condition twice, separated by a one month interval. The ordering of the experimental conditions for the first session was reversed for the second session (seeFigure 7).

To ensure that an equal number of data points were collected from each console, 10 simulated tracks per hour were randomly presented in each of the 2 (or 4) geographic zones — unlike in Experiment 1, where a constant 25 tracks per hour were randomly presented across all zones. This resulted in 20 simulated tracks per hour presented during the 2 zone condition and 40 simulated tracks per hour during the 4 zone condition. In total, 3840 simulated tracks were injected during the entire experiment. In contrast to Experiment 1 all simulated tracks were of the same type: search tracks. NORAD officials consider detection of all search tracks their primary task because those not identified may pose a threat to North American security. Also, since RTs to beacon tracks were relatively fast and stable throughout Experiment 1 the decision was made to concentrate on search tracks

	Evening Shift	Midnight Shift
Day 1	2 zones 20/20	2 zones 20/20
Day 2	4 zones 20/20	4 zones 20/20
Day 3	2 zones 60/60	2 zones 60/60
Day 4	4 zones 60/60	4 zones 60/60

Session 1

1 month interval

	Evening Shift	Midnight Shift
Day 1	4 zones 60/60	4 zones 60/60
Day 2	2 zones 60/60	2 zones 60/60
Day 3	4 zones 20/20	4 zones 20/20
Day 4	2 zones 20/20	2 zones 20/20

Session 2

Figure 7: Presentation order of treatment conditions in experiment 2.

3.2 Results

A total of 3322 out of a possible 3840 data points (86.5%) were collected during the experiment — the remainder (518) were lost due to computer problems. Of the collected data, approximately equal numbers were retrieved from both experimental sessions (41.5% for session 1, 45% from session 2). The strategies developed in Experiment 1 for normalizing the data and identifying outliers were applied to these data and resulted in a further 19 data points being eliminated. As in Experiment l, all statistical analyses were performed on the transformed data and group differences are reported as geometric means with 5-95% confidence intervals.

A 2x2x2x2 ANOVA was performed on the following four factors: shift (evening vs. midnight); number of geographic zones (2 vs. 4); time-on-task (20/20 vs. 60/60); and work-session (session 1 vs. session 2).There was a statistically significant main effect for number of zones ($F(1,15) = 81.3$, $P<.0001$). When the surveillance area of Canada East was divided into 4 zones the mean RT was 74.9 sec (73.8 - 76.1 sec). For the 2 zone condition the mean RT was 20 seconds slower — 94.0 sec (91.9 - 96.1 sec). This poorer performance is likely due to the smaller visual angles the targets subtended in the 2 zone condition, making detection more difficult. Parenthetically, this result surprised the surveillance operators who, almost unanimously, told the experimenters (during the study) that two surveillance operators could perform the task as efficiently as four.

There was no main effect for time-on-task, but there was a shift by time-on-task interaction ($F(1,15) = 7.3$, $P<.0164$). The 60 min work, 60 min rest schedule induced (≈10 sec) longer RTs for subjects working the midnight shift than did the 20/20 work/rest schedule (see Figure 8). This suggests that only the midnight shift was susceptible to a vigilance decrement.

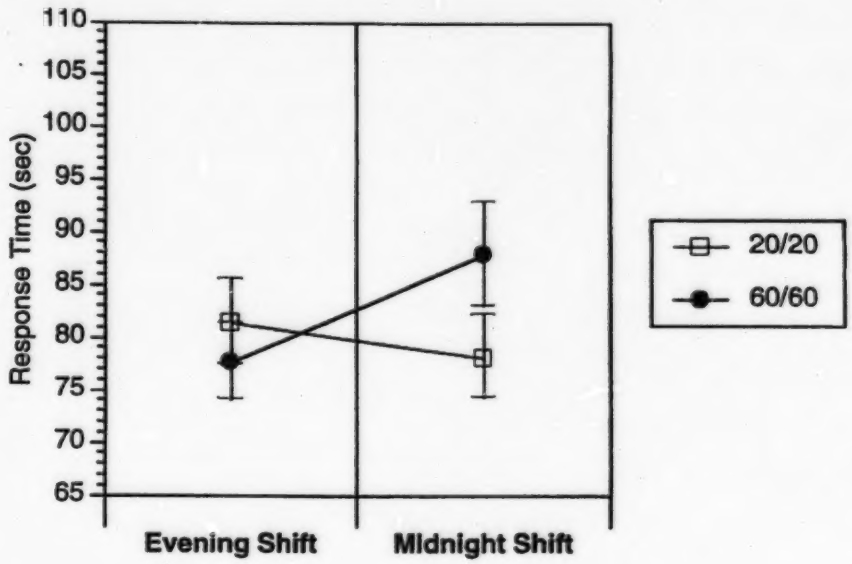

Figure 8: Reponse times for the shift by time-on-task (work/rest schedule) interaction.

In order to investigate the source of this vigilance decrement, RTs were plotted for the evening and midnight shifts for each of the four experimental conditions (see Figure 9). As was seen earlier, the 2 zone condition yielded (overall) longer RTs than the 4 zone condition. However, a statistically significant simple effect was observed only between the 20/20 and 60/60 conditions during the midnight shift for 4 zones. It seems that subjects demonstrated a vigilance decrement only when the following two conditions were satisfied: 1) it was the midnight shift, and 2) Canada East was divided into 4 rather than 2 geographic zones. It is interesting to note that this vigilance decrement was modulated by circadian influences. Figures 10a & b depict hourly RT performances for each work/rest schedule in the 4 zone condition for both the evening and midnight shifts. When subjects are given a 20 min work, 20 min rest schedule hourly performance was relatively stable throughout both shifts. During the 60 min work, 60 min rest schedule, however, a gradual deterioration in performance occurred during the period between 3 and 6 AM of the midnight shift — an effect similar to that observed in Experiment 1.

Also consistent with the first experiment was an improvement in performance after the beginning of a shift. As mentioned earlier, this result could be due to sociological and/or psychological factors. During Experiment 2 we noticed higher levels of non-task related communication among the surveillance operators at the beginning of each shift — suggesting social influences. However, a psychological factor may also have contributed to the effect. Figure 11 depicts changes in RT in 10 minute intervals within the 20/20 and 60/60 work/rest conditions. There was a dramatic (≈ 24 sec) improvement in performance from the first to the second 10 minutes in the 60/60, 2 zone conditions. It seems that, in the more difficult 2 zone condition after 60 minutes of rest, the surveillance operators needed 10 minutes of 're-adaptation' before reaching optimal performance. Although a similar trend exists for the 20/20 condition, it is not statistically significant. Also, these results suggest that there was no within work period vigilance decrement. In both the 2 and 4 zones 60/60 conditions, performance remained stable — i.e., did not degrade — after the first 20 minutes.

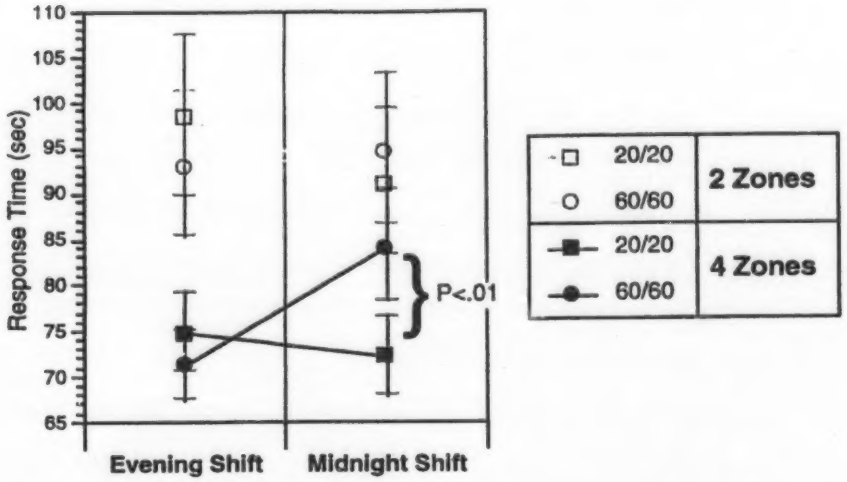

Figure 9: Evening and midnight shift response times for the coverage (# of zones) and time-on-task conditions.

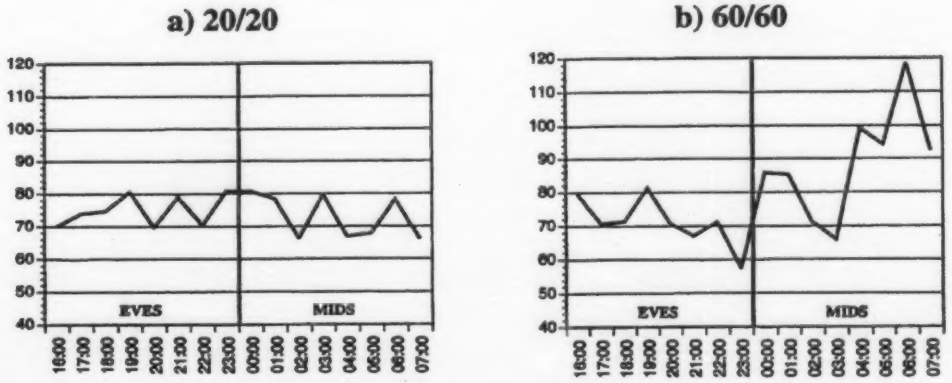

Figure 10: Hourly response times for the evening and midnight shifts in the 4 zones condition: a) 20/20 work/rest schedule; b) 60/60 work/rest schedule.

Why did a vigilance decrement manifest itself for the midnight shift in the 4 zone condition, but not in the more difficult 2 zone condition? To answer this question the data were analysed with respect to geographic zone. Figures 12a & b illustrate RTs by work/rest schedule and shift for the 2 zone condition. There were no statistically significant differences in RT between zones for either the 20/20 or 60/60 work/rest (time-on-task) conditions. This is despite the fact that a large difference exists between the proportion of air traffic traversing the East (E) vs. the North (N) zone. These results are contrasted with those presented in Figures 13a & b in which there were statistically significant shift by time-on-task ($F(1,15) = 11.0$, $P<.0047$) and shift by zones interactions ($F(3,15) = 11.2$, $P<.0004$). Figure 13a shows that operators working the midnight shift in the 20/20 work/rest condition performed more poorly from the South East (SE) to the northern zones, whereas performance across regions for the evening shift remained stable. The vigilance decrement during the midnight

shift is even more pronounced in the 60/60 work/rest condition, particularly for the East (E) and North East regions (Figure 13b). These results suggest that operators working northern regions (i.e., low air traffic) during the midnight shift were more susceptible to vigilance decrements, particularly during the tiring 60/60 work/rest condition. Although there is certainly a circadian influence contributing to these results, this explanation does not account for the unchanging performance observed in the 2 zone condition.

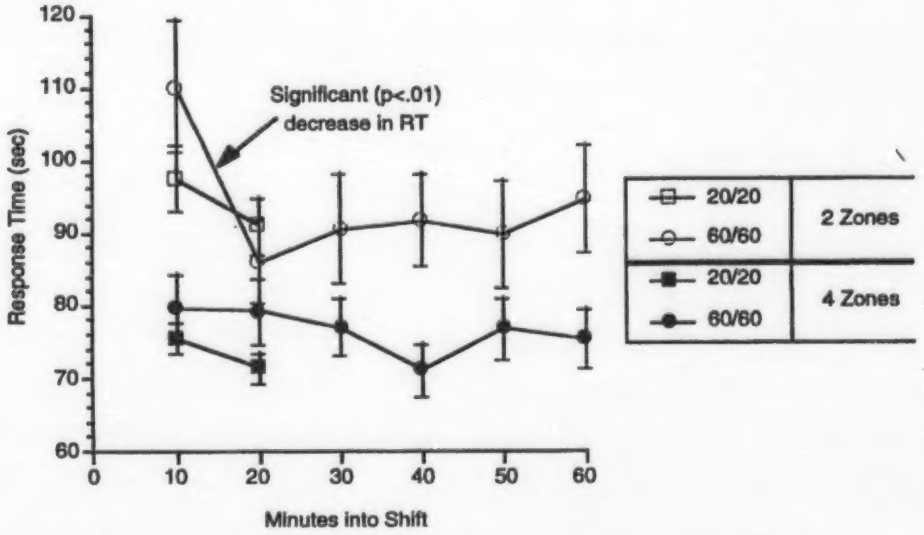

Figure 11: Response times in 10 min intervals from the begining of each work period.

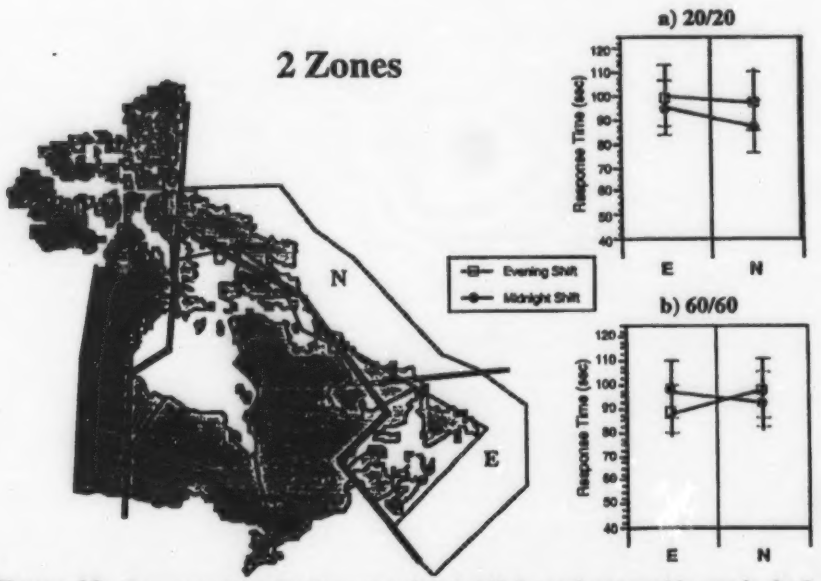

Figure 12: Response times for time-on-task and shift for each geographic area in the 2 zones condition.

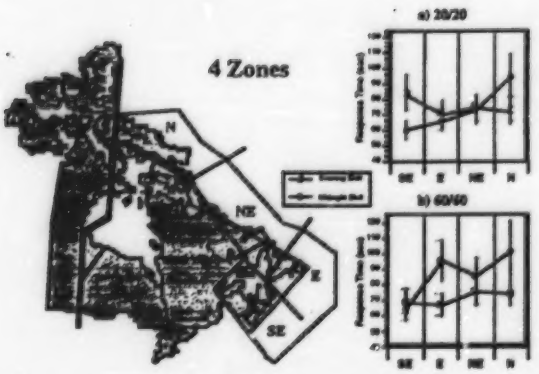

Figure 13: Response times for time-on-task and shift for each geographic area in the 4 zones condition.

It is unclear why the more difficult 2 zone condition did not show a similar decline in performance from the East to the North zones. We expect that motivation may have played a key role. During the experiment (4 zone condition) we noticed that operators in the midnight shift, surveying the northern zones, attended to their consoles less and communicated with their neighbors more. This may have been an unconscious strategy among the operators to keep themselves awake while monitoring low (i.e., boring) traffic areas. In contrast, the northern most operator in the 2 zone condition was probably discouraged to engage in idle discussion because his/her neighbor was busy with the more difficult East zone; hence, perhaps reinforcing better performance in the northern operator's area. Another possible motivational influence was the physical proximity to the surveillance supervisor in the 2 zone condition. In NORAD, the role of the surveillance supervisor (who did not take active part in the experiment) is to ensure the efficient operation of the surveillance team. Figures 14a & b illustrate the placement of the consoles with respect to the supervisor's console and show that, unlike the northern most consoles in the 4 zone condition, consoles in the 2 zone condition were situated near the supervisor. It is possible that this proximity to the supervisor could have provided extra motivation — motivation necessary to overcome boredom and fatigue from surveying the north during early morning hours.

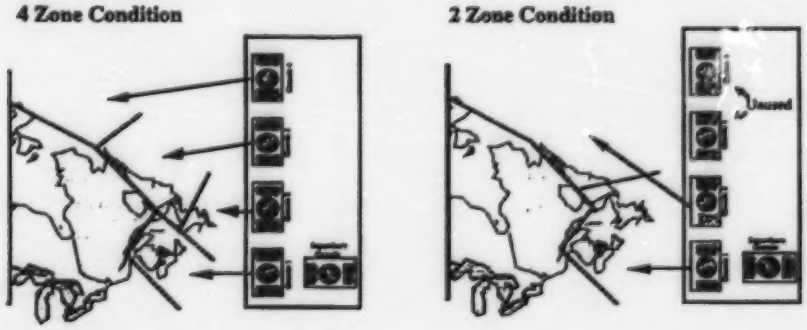

Figure 14: Physical layout of consoles for the 2 and 4 zones conditions.

There was no main effect for work-session, indicating there was (overall) no difference between the first and second testing sessions. However, there were 2 two-way interactions: work-session by shift ($F(1,15) = 27.3$, $P<.0001$) and work-session by time-on-task ($F(1,15) = 18.1$, $P<.0007$). Figure 15a shows that between the first and second testing periods RT increased during the midnight shift but decreased during the evening shift. This decrease in response times for the evening shift is likely due to differential radar clutter. During the first two days of session 1 there was considerable radar clutter in the South and South-East zones. Relatively little clutter was present for all geographic zones during all of session 2. Although the midnight shift similarly experienced clutter during the first two days of session 1, RTs increased rather than decreased in session 2. Ironically, a simpler detection task for the midnight shift may have had a de-arousal effect on the subjects making them more susceptible to circadian influences. The results in Figure 15b, which display the work-session by time-on-task interaction, support this interpretation. From session 1 to session 2 RT increased for the 60/60 work/rest condition but slightly decreased for the 20/20 condition.

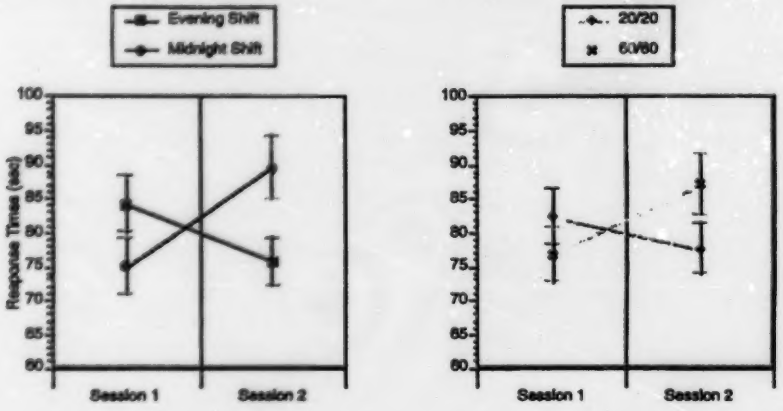

Figure 15: Experimental session by shift (a) for time-on-task (b) and interactions.

In summary, Experiment 2 yielded the following findings: 1) detection times took approximately 20 seconds longer when Canada East was divided into 2 rather than 4 geographic zones; 2) the 60/60 work/rest condition produced an overall vigilance decrement only for subjects working the midnight shift in the 4 zone condition; 3) in the 4 zone condition for the midnight shift, operators surveying northern zones demonstrated greater vigilance deficiencies than those surveying southern zones; 4) these subjects were also susceptible to circadian influences; and 5) a (≈ 10 min) period of readjustment was needed for subjects to reacquaint themselves with the detection task in the 2 zone, 60/60 condition.

3.3 Discussion

The results from this study are consistent with Parasuraman's [16] assertion that, "...vigilance failures in complex monitoring do occur and may result either from a vigilance decrement or from a low overall level of vigilance" (p. 700). Contrary to other researchers [e.g., 1, 13] Parasuraman [15, 16] has maintained that the vigilance decrement is a robust effect and can be found in real world situations. Although our results suggest that, in the NORAD environment, vigilance failures do occur, the lack of an overall vigilance decrement (i.e., main effect for time-on-task) in Experiment 2 is also consistent with criticisms that the effect generalizes poorly from the laboratory to the real world [1, 11, 12, 13]. The 60/60 vs. 20/20 work/rest condition produced longer response times

only under conditions where the geographic region was divided into 4 zones and the surveillance operators worked midnight shifts. As in Experiment 1 the gradual decline in vigilance during early morning hours is consistent with chronopsychological studies showing that cognitive performance is sensitive to circadian influences e.g., [3, 8, 9].

Among the many factors posited in the literature to explain small or nonexistent vigilance decrements in real world settings are target probability, stimulus rate, stimulus complexity, task type, skill level and motivation. It has been suggested that a stimulus rate of at least 24 [14] or 20 [23] events per minute is needed to observe a vigilance decrement. The stimulus rate (i.e., target + non-target) in our study could not be precisely calculated because the number of non-targets (i.e., spurious radar tracks) varied dynamically with the physical environment. Nevertheless, we estimate that, given the relatively large amount of radar 'clutter' present on the operators' consoles during the experiments, a stimulus rate at least as great as 24 events/min was occurring. Even though the number of simulated tracks per consoles was kept constant, the overall target rate (i.e., simulated + live tracks) varied from ≈ .2 event/min in relatively sparse northern zones to ≈ 1 event/min in high traffic areas. There are two reasons why it is unlikely that this difference in target probability is responsible for the poorer performance observed among operators working the northern zones. First, there was no similar effect for the 2 zone condition; and second, the 4 zone/evening shift condition also showed no effect despite comparable target probability distributions among the respective geographic zones.

Parasuraman [14] distinguished between two types of vigilance tasks: (1) those that involve successive-discrimination where "...the target and nontarget features are not present at the same time" (p. 924) and thus induces short term memory loads; and (2) simultaneous-discrimination tasks where targets are at all times fully discriminable from nontargets. In Experiment 1, the detection of beacon tracks, with their unique symbology, could be classified as a simultaneous-discrimination task. The detection of search tracks, on the other hand, involves successive-discrimination because targets can be distinguished from nontargets only by comparing 'blip' trail lengths. From a survey of the literature Parasuraman [14] established that reliable vigilance decrements are observed only for successive-discrimination tasks with high event rates. Our results are consistent with this conclusion. Response times for detecting beacon tracks in Experiment 1 did not indicate vigilance failures, whereas response times for search tracks did. This factor alone, however, does not explain why an overall vigilance decrement — main effect for time-on-task — failed to occur in Experiment 2 where all stimuli were of the successive-discrimination type.

NORAD personnel are highly trained and skilled surveillance operators. It is possible that this training and experience could have offset any vigilance decrement expected from a 60/60 work/rest schedule. Parasuraman and Giambra [15] have shown that practice can reduce or even eliminate vigilance decrements. In their study, three groups of subjects (young = 19-27 yrs, middle aged = 40-55 yrs. and old = 70-80) were allowed 20 practice sessions on a 30 minute vigilance task. During the first two sessions the percentage hit rate decreased for all subjects from the first to the third 10 minutes of the task — i.e., a noticeable vigilance decrement. By the 19th and 20th practice sessions, all three age groups showed marked improvements in resisting vigilance decrements with the youngest group totally eliminating it. The subjects in our study were all young and well trained military personnel. This combination of youth and extensive practice could account for the lack of an overall decrement effect.

A vigilance decrement occurred only for operators working the midnight shift when Canada East was divided into 4 surveillance zones. It is known that vigilance decrements are sensitive to diurnal influences [5, 18], but it is unclear why the effect was not exhibited for the 2 zone condition. We posit that motivation played an important role to assuage circadian induced vigilance decrements. Warm, Dember, Lanzetta, Bowers, and Lysaght [21] demonstrated that by increasing the level of complexity of a task, vigilance decrements can be reversed. They attributed this finding to greater intrinsic motivation elicited by tasks of greater complexity. The 2 zone condition elicited response times 20 sec longer (overall) than those observed during the

4 zone condition. This suggests that the surveillance task was more difficult to perform and, as such, could have provided greater intrinsic motivation to maintain efficiency during the midnight shift.

Intrinsic motivation is motivation elicited by an inherent interest in the task. Extrinsic motivation — i.e., pressure to perform imposed by an outside source — may also have played a role in the results. Horne and Pettitt [10] found that subjects performing an auditory vigilance task can resist circadian induced vigilance decrements if sufficient monetary incentives are given. The ameliorating effects of this extrinsic motivation did not continue, however, past the third night of sleep loss in their study. In the NORAD environment, the surveillance supervisor is responsible for ensuring that all radar tracks entering Canadian air space are quickly detected by the team of surveillance operators. Consequently, this person can be a source of considerable influence and motivation to the subjects. Simple physical proximity to the supervisor may motivate operators to maintain performance during the early morning hours. Both surveillance operators were situated near the supervisor in the 2 zone condition. In the 4 zone condition, operators of northern zones were situated further away. From informal observations during the experiment, we noticed that northern operators were allowed greater latitude (by the supervisor) to partake in non-task related behaviour. As a result, these operators may have experienced less extrinsic motivation to efficiently perform their task. We suggest therefore that, among these highly trained individuals, a vigilance decrement is manifested only when the following two conditions are met: 1) low central nervous system arousal due to circadian influences (i.e., midnight shift), and 2) reduced intrinsic and/or extrinsic motivation to properly perform the task (i.e., less demanding task and/or less supervision).

Our finding that performance improves from the first to the second 10 minutes on task, after a 60 minute break for the 2 zone condition, is intriguing. For this real world task, it appears that a break lasting 60 minutes is sufficient to require a period of cognitive/perceptual readjustment after which performance improves (a vigilance increment effect?). Recall that the subjects rotated among the consoles in both the 20/20 and 60/60 work/rest conditions, thus encountering different radar information each work period. The effect is therefore not due to the radar picture simply having had more time to change during the 60 vs. 20 min rest period. It appears that highly skilled surveillance operators need approximately 10 minutes of 'practice', after a 60 min break, to reach asymptotic levels of performance in the more complex and less familiar 2 zone condition. This interpretation is consistent with Pellegrino, Doane, Fischer and Alderton's [17] study in which they found that improvements in performance to a perceptual comparison task was a function of both the amount of practice and the context in which responses were elicited. It appears that in the more familiar and less difficult 4 zone condition, a period of 'practice' is not necessary to regain optimal performance after a 60 min break.

In conclusion, our results suggest that a vigilance decrement effect can occur in a real world detection task. The effect, however, is 1) not nearly as strong as those reported in laboratory studies and 2) is limited largely to the midnight shift. Also, with the judicious use of intrinsic and/or extrinsic motivation, it may be possible to eliminate this vigilance decrement effect altogether.

4 ACKNOWLEDGEMENT

The authors would like to thank the Commander and staff of the 22 Wing at Canadian Forces Base North Bay for their assistance with these studies. In particular we would like to thank the crew of 21 Aerospace Warning and Control Squadron, who were subjects in the experiments, as we as the 'sim' input team for their efforts. A working group, headed by LCol J.A. Nelson, deserves thanks for implementing th experiment. Among this team, two people deserve special mention: Warrant Officer W.E. Horsman who organized the sim inputs and Captain M.Y. Dubois who undertook the data reduction.

5. REFERENCES

1. Adams, J. A. (1987). Criticisms of vigilance research: A discussion. Human Factors, 29 (6), 737-740.

2. Alluisi, E. A., Coates, G. D. and Morgan, B. B. J. (1977). Effects of temporal stressors on vigilance and information processing. In R. R. Mackie (Ed.), Vigilance: Theory, operational performance, and physiological correlates. (pp. 361-421). New York: Plenum Publishing Company.

3. Angus, R. G., Pigeau, R. A. and Heslegrave, R. J. (1992). Sustained-operations studies: From the field to the laboratory. In C. Stampi (Ed.), Why We Nap: Evolution, Chronobiology, and Functions of Polyphasic and Ultrashort sleep. (pp. 217-241). Boston: Birkhauser.

4. Bradfield, A.P., O'Neill, P.F., Pigeau, R. A. and Angus, R. G. (1989). Results of a preliminary study of Air Surviellance Operator performance. (DOR Staff Note 5/89) Fighter Group, Canadian NORAD Region Headquarters, North Bay, Ontario.

5. Craig, A., Davies, D. R. and Matthews, G. (1987). Diurnal viariation, task characteristics, and vigilance performance. Human Factors, 26 (6), 675-684.

6. Davies, D.R., & Parasuraman, R. (1982). Effects of the Environment and General State on Vigilance. In The Psychology of Vigilance (pp. 141-179). London: Academic Press.

7. Elliott, E. (1960). Perception and alertness. Ergonomics, 3 357-364.

8. Haslam, D. R. (1982). Sleep loss, recovery sleep and military performance. Ergonomics, 25 163-178.

9. Horne, J. A. (1988). Why we sleep: The functions of sleep in humans and other mammals. Oxford: Oxford University Press.

10. Horne, J. A. and Pettitt, A. N. (1985). High incentive effects on vigilance performance during 72 hours of total sleep deprivation. Acta Psychologica, 58 123-139.

11. Kibler, A. W. (1965). The relevance of vigilance research to Aerospace monitoring tasks. Human Factors, 7 (2), 93-99.

12. Mackie, R. R. (1984). Research relevance and the information glut. In F. A. Muckler (Ed.), Human Factors Review, (pp. 1-11). Santa Monica, CA: Human Factors Society.

13. Mackie, R. R. (1987). Vigilance research -- are we ready for countermeasures? Human Factors, 29 (6), 707-723.

14. Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. Science, 205 924-927.

15. Parasuraman, R. and Giambra, L. (1991). Skill development in vigilance: effects of event rate and age. Psychology of Aging, 6 (2), 155-169.

16. Parasuraman, R., Warm, J. S. and Dember, W. N. (1987). Vigilance: Taxonomy and utility. In L. Mark, J. S. Warm and R. L. Huston (Eds.), Ergonomics and Human Factors: Recent Research, (pp. 11-32). New York: Springer-Verlag.

17. Pellegrino, J. W., Doane, S. M., Fischer, S. C. and Alderton, D. (1991). Stimulus complexity effects in visual comparisons: The effects of practice and learning context. Journal of Experimental Psychology: Human Perception and Performance, 17 (3), 781-791.

18. Riemersma, J. B. J., Sanders, A. F., Wildervanck, C. and Gaillard, A. W. (1977). Performance decrement during prolonged night

driving. In: Vigilance: Theory, Operational Performance, and Physiological Correlates. Proceedings of a Symposium on Vigilance II: Relationships among theory, physiological correlates, and operational performance (pp.41-58). St. Vincent, Italy: Plenum Press in coordination with NATO Scientific Affairs Division.

19. Sokal, R. R., & Rohlf, F. J. (1969). The logarithmic transformation. In R. Emerson, D. Kennedy, R. B. Park, G. W. Beadle, & M. Whitaker (Eds.), Biometry (pp. 382-385). San Francisco: W.H. Freeman and Company.

20. Thackray, R. I., Bailey, J. P. and Touchstone, R. M. (1977). The effect of increased monitoring load on vigilance performance using a simulated radar display (FAA-AM-77-18). FAA Civil Aeromedical Institute.

21. Warm, J. S., Dember, W. N., Lanzetta, T. M., Bowers, J. C. and Lysaght, R. J. (1985). Information processing in vigilance performance: Complexity revisited. In: Proceedings of the 29th Annual Meeting of the Human Factors Society (pp.19-23). Santa Monica, CA:

22. Wiener, E. I. (1987). Application of vigilance research: Rare, Medium, or Well Done? Human Factors, 29 (6), 725-736.

23. Williams, P. S. (1986). Processing demands, training and the vigilance decrement. Human Factors, 28 (5), 567-579.

#506689